USAID
FROM THE AMERICAN PEOPLE

المملكة المغربية
وزارة التربية الوطنية
والتكوين المهني

USAID/MOROCCO READING FOR SUCCESS SMALL-SCALE EXPERIMENTATION ACTIVITY (RFS-SSE)

EARLY GRADE READING ASSESSMENT MIDLINE 1 REPORT

Contract No. AID-608-TO-15-00002

Submission Date: 5 September 2016

EARLY GRADE READING ASSESSMENT

MIDLINE 1 REPORT

RFS-SSE

CONTRACTED UNDER ORDER NO. AID-608-TO-15-00002

ACKNOWLEDGEMENTS

The School-to-School International and Chemonics International team would like to thank USAID/Morocco for their support. Particularly, we wish to acknowledge the important contribution of Thomas LeBlanc and Mariam Britel-Swift from the education office at USAID/Morocco. The USAID education office has provided guidance and support throughout the preparation and administration of the study.

We would also like to recognize the staff from the Ministry of Education and Vocational Training (MOE) and namely the National Center for Evaluations and Exams (CNEE) led by Mr. Mohammed Sassi, Director of the Center, Mr. Ahmed Chaibi, head of the Evaluation Division, Mr. Said Bouderga, head of the Examinations Division, and Mrs. Hafida Kach, manager at CNEE. The CNEE has played an important role at all stages of the study: sampling, tool development, identifying data collectors, development of the data analysis plan, and validation of the Midline 1 findings.

We would also like to thank the Director of the Curricula Directorate (Direction des Curricula), Mr. Fouad Chafiqi, and his team for their continuous support in the implementation of the RFS-SSE activity.

We also extend our thanks to the regional and provincial offices (AREFs), which provided all the necessary field support through the mobilization of school inspectors who have effectively contributed to the development of the EGRA test and SSME questionnaires as well as inspectors and counselors who led data collection with great professionalism.

The following persons played a major role in the process:

Abdelaziz El Ghordaf	(Chemonics International)
Selma Talha Jebril	(Chemonics International)
Shamineh Byramji	(Chemonics International)
Rida Baidas	(Chemonics International)
Maryvonne Denauw	(School-to-School International)
Beth Odenwald	(School-to-School International)
Selene Rangel	(School-to-School International)
Dr. Michel Rousseau	(School-to-School International)
Claire Wasserman	(School-to-School International)

CONTENTS

LIST OF TABLES AND FIGURES

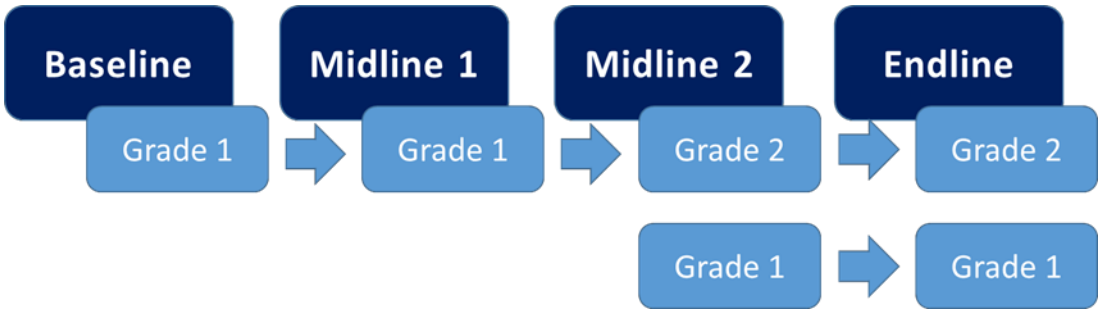
ACRONYMS

AREF	Académie Régionale d'Education et de Formation
CFRN	Centre Des Formations et Des Rencontres Nationales (Center for Trainings and National Assemblies)
CNEE	National Center for Evaluations and Exams
CSO	Civil Society Organizations
CWPM	Correct Words Per Minute
EDO	Education District Officers
EGRA	Early Grade Reading Assessment
EMIS	Education Management Information Systems
MOE	Ministry of Education and Vocational Training
MSA	Modern Standard Arabic
ORF	Oral Reading Fluency
RFS-SSE	Reading for Success-Small Scale Experimentation
SSME	Snapshot of School Management Effectiveness
STS	School-to-School International
USAID	United States Agency for International Development

EXECUTIVE SUMMARY

This report presents key findings of the Morocco **Reading for Success - Small Scale Experimentation (RFS-SSE)** Early Grade Reading Assessment (EGRA) Midline 1. The overall intent of the RFS-SSE evaluation is to assess improvement in reading among Grade 1 and Grade 2 students in 90 schools across eight delegations in Morocco. This evaluation tests students in 45 intervention schools (experimental group) and 45 non-intervention schools (control group). There will be four evaluations conducted at intermittent points throughout the life of the project (see image below).

FIGURE 1. STUDY DESIGN BY EVALUATION CYCLE



The Baseline was conducted in January 2016 and the first of two midlines, hereafter referred to as "Midline 1", was conducted in May 2016. Midline 2 is scheduled for September 2016 to coincide with the commencement of the new academic year. The Baseline captured the learning levels of students in the middle of their Grade 1 year, immediately prior to the implementation of RFS-SSE lesson plans. The Midline 1 captured the learning levels of those same students at the end of Grade 1. The Midline 2 will again measure these same students from the previous year, now Grade 2 students, as well as a new group of students just admitted into Grade 1 for the 2016-2017 school year. The Endline for both grades will be conducted at the end of the academic year in May 2017. When reviewing the results of the Midline 1, it is important to keep in mind that the time period between the Baseline (January 2016) and Midline 1 (May 2016) was very short. Therefore, we should not expect to see great differences in scores between the two time points.

The table below summarizes the EGRA results from the Baseline and Midline 1. The table summarizes the mean scores of the experimental and control groups for each of the six subtasks that were included in the EGRA.

Average Scores for all EGRA Subtasks

EGRA Subtask	Experimental group		Control group		Experimental group	Control group	Difference
	Baseline	Midline 1	Baseline	Midline 1	Change (x)	Change (y)	x - y
Phonemic Awareness	3.11	6.19	3.33	4.89	3.08	1.56	1.52
Syllable Identification	23.27	37.60	23.13	37.09	14.33	13.96	0.37
Nonword Reading	7.20	11.41	6.86	11.52	4.21	4.66	-0.45
Passage Reading	5.83	13.28	5.59	13.04	7.45	7.45	0.00
Reading Comprehension	0.36	0.31	0.38	0.36	-0.05	-0.02	-.03
Listening Comprehension	1.13	1.75	1.15	1.59	0.62	0.44	0.18

Of the six subtasks assessed at Baseline and Midline 1, two showed a statistically significant difference between the experimental and control groups in the progress made since Baseline: **Phonemic Awareness** and **Listening Comprehension**.

Comparing the results of Baseline to Midline 1, we observed that students from the experimental group were, on average, able to recognize 3.08 more phonemes at Midline 1 compared to Baseline, while the students from the control group were, on average, able to recognize only 1.56 more phonemes at Midline 1 compared to Baseline. This difference between the progress of the two groups is statistically significant ($t(1675) = -5.33$; $p<0.000$).

RFS-SSE's technical approach is to introduce phonemic awareness and a focus on phonics into the curricula, as we know these fundamentals are building blocks on which all other aspects of reading competency is built: deciphering words, fluency, and comprehension. Phonological and phonemic awareness are often strong predictors of reading achievement. Therefore, a significant increase in the phonemic awareness subtask between the experimental and control groups indicates the start of a solid reading foundation among the experimental group. Our expectation is that this foundational phonemic awareness will provide the experimental group with a greater ability to improve performance in other reading competencies, after given sufficient time. We will have the opportunity to see how the experimental group improves on other subtasks during the remaining two evaluations, in September 2016 and May 2017.

There was also a statistically significant difference between the experimental and control groups in the progress made since Baseline on the Listening Comprehension subtask. The number of correct answers increased since Baseline, on average, more in the experimental group than in the control group. Students from the experimental group provided, on average, 0.62 more correct answers at Midline 1 compared to Baseline while those from the control group provided, on average, only 0.44 more correct answers at Midline 1 compared to Baseline. The difference between the two groups is statistically significant ($t(1675) = 2.36$, $p=0.021$).

The table below summarizes the proportion of zero scores from the Baseline and Midline 1. The table summarizes the percent of students from the experimental and control groups that received a zero score for each of the six subtasks that were included in the EGRA.

Percent of Zero Score Students for all EGRA Subtasks

EGRA Subtask	Experimental Group		Control group		Experimental group	Control group	Difference
	Baseline	Midline 1	Baseline	Midline 1	Change (x)	Change (y)	x - y
Phonemic Awareness	51.73	22.70	48.93	35.60	29.03	13.33	15.70
Syllable Identification	15.89	8.48	16.43	8.33	7.41	8.10	-0.69
Nonword Reading	39.78	24.49	41.31	28.10	15.29	13.21	2.1
Passage Reading	45.28	21.62	45.71	25.12	23.66	20.59	3.07
Reading Comprehension	75.39	77.42	76.67	76.67	-2.03	0.00	-2.03
Listening Comprehension	45.64	28.32	46.43	31.31	17.32	15.12	2.20

Of the six EGRA subtasks, there was a statistically significant difference observed in the percent of zero score students among the experimental and control groups for the **Phonemic Awareness** subtask only. In the control group, 35.60% (n=299) of the students received a zero score (i.e., had 0 correct answers) on this subtask at Midline 1, while in the experimental group, only 22.70% (n=190) of the

students received a zero score. The difference between the two groups is statistically significant (χ^2 (1,677) = 33.75, $p<0.01$). The

No other statistically significant differences between the experimental and control groups were observed in the progress made since Baseline on students' EGRA scores. This is likely due to the limited amount of time that the intervention was active in classrooms in between the Baseline in January 2016 and the Midline 1 in May 2016. The intervention was only implemented for two and a half months during the second semester of Grade 1. This limited timeframe was compounded by schools having a two-week vacation during the month of April and having to start the data collection in mid-May, immediately prior to the end of the school year. The remaining two evaluations to be conducted in September 2016 and May 2017 will allow us to see if the initial trends observed at Midline 1 continue throughout the following school year as students' exposure to the RFS-SSE intervention increases.

CHAPTER 1: INTRODUCTION

1.1 Reading for Success – Small-Scale Experimentation (RFS-SSE)

The Reading for Success – Small-Scale Experimentation (RFS-SSE) activity is the current active component of a broader USAID initiative. RFS-SSE is designed to reflect ongoing collaborations between USAID/Morocco and the Ministry of Education and Vocational Training (MOE) to improve reading instruction in Morocco. Conceived as a learning activity, RFS-SSE will develop an evidence base of effective approaches that can improve reading skills in targeted primary schools. RFS-SSE comes at a time when the MOE is developing a 15-year education reform called Vision 2030 as well as a set of medium-term activities for the period 2015-2020. Reform efforts will address a key weakness in the Moroccan educational system: poor reading skills at the primary level. The RFS-SSE intervention will help to inform the revisions to the existing curriculum and the design of the reformed curriculum by providing data and evidence to support the envisioned changes.

RFS-SSE is designed to test activities and to strengthen stakeholder engagement to improve reading skills in the early grades. It is the second phase of the larger RFS process that spans from analysis to national implementation of the Reading for Success (RFS) Program. RFS-SSE is using previous research and related activities carried out in Morocco and other countries and drawing from lessons learned to test two reading interventions: 1) a new approach to teaching reading in Arabic and 2) summer enrichment activities. The new Arabic reading lessons are based on a phonetic approach, where each lesson includes a focus on phonics and syllabic sounds in addition to a focus on vocabulary acquisition through oral story reading and discussion.

Reading lessons are being developed and tested and results will feed into planned MOE curriculum reforms linked to the medium-term activities for 2015-2020. These efforts have been undertaken to better understand the technical assistance needed for implementing activities that improve reading instruction in government primary schools and to test the effectiveness of new reading materials. Similarly, the project will engage civil society organizations (CSOs) during the second summer of the project and support and test their efforts to reduce learning loss over the summer months through enrichment activities.

1.2 Context

RFE-SSE has developed Modern Standard Arabic (MSA) reading lessons based on a phonetic approach combined with vocabulary enhancement techniques through stories that will improve reading comprehension and communication for primary school students in Grades 1 and 2. These new reading lessons will be introduced in 90 schools across eight delegations in four regions of Morocco.

The new approach for Grade 1 is based on the phonetic/syllabic teaching method, where each reading lesson is comprised of two axes: reading acquisition based on phonetic syllabic sound (phonemic awareness) and vocabulary acquisition/enrichment through stories and thematic texts. The phonemic-based approach teaches students to learn that words are made up of sounds which can be put together in different combinations to make a variety of words. By introducing stories into the reading lessons, students are acquainting themselves with new words and building their vocabulary while making the learning interactive, fun, and collaborative for both the students and their teachers.

To date, RFS-SSE has focused the reading intervention in schools during the second semester of the school year for Grade 1 students only. Therefore, the Baseline conducted in January 2016 and the Midline 1 conducted in May 2016 assessed only Grade 1 students. During the 2016-2017 academic year, the intervention will be implemented in both Grades 1 and 2. Therefore, the Midline 2 (in September 2016) and the Endline (in May 2017) will assess students in both Grades 1 and 2.

As part of the curriculum development, RFS-SSE developed lesson plans for teachers that guide them through their daily reading lesson. These lesson plans were incorporated into a teacher's guide that provides the teachers with background on the approach and suggested techniques. In order to apply the vocabulary enrichment track in the classroom, each student was provided a storybook containing three stories that were read during the semester. The teachers in the intervention schools were trained on the new approach by RFS-SSE trainers.

Following the first semester of implementation, the curricula and lesson plans are currently being revised for Grade 1 and being developed for Grade 2, prior to the start of the new academic year, 2016-2017.

1.3 Evaluation Design

In order to assess the impact of the RFS-SSE reading program, a longitudinal evaluation design (including equal representation of girls and boys in both urban and rural schools) was selected.

As shown in **Error! Reference source not found.** in the Executive Summary, the evaluation design includes reading assessments of two cohorts of students. The first cohort of students is assessed at four time points – the middle of Grade 1 and throughout Grade 2 (Baseline, Midline 1, Midline 2 and Endline). The second cohort of students is assessed at two time points – the beginning and end of Grade 1 (Midline 2 and Endline). Please see image in the executive summary.

At Baseline, all students were assessed with the same EGRA instrument that was developed to test students at either a Grade 1 or Grade 2 level. Because the students were selected at random, it is possible to generalize the results to all students in the treatment group in the eight delegations where the reading lessons are being implemented.

1.4 What are EGRA and SSME?

The Early Grade Reading Assessment (EGRA) was developed in consultation with cognitive scientists, early grade reading experts, research methodologists, and assessment experts with funding assistance from USAID,the World Bank, and other international donors . This assessment tool measures student performance on the basic foundational skills required for fluency in reading. EGRA assesses the skills needed for reading acquisition. Although many students are not yet fluent readers in the early grades, the EGRA allows us to capture what students, even the "nonreaders", *can* do and where they are in the developmental path to becoming fluent readers.

EGRA is an individually administered, oral assessment that requires approximately 20 minutes per student. The test is administered to one student at a time by a trained enumerator (not teachers) in a location outside of the classroom. The enumerator begins by explaining the assessment to the student and asking if the student agrees to participate. Consent is always optional and no student is required to take the assessment. The enumerator creates a relaxed environment for the student and assures the student that the assessment is not used for a grade. The enumerator then begins by asking the student questions aloud and having the student respond aloud. For defined subtasks, the enumerator places a paper stimulus in front of the student containing letters or words, and the enumerator asks the student specific questions about the stimuli. The EGRA administered for the RFS-SSE Midline 1 consisted of six subtasks which are described in more detail below.

The Snapshot of School Management Effectiveness (SSME) is an instrument that captures a picture (i.e., "snapshot") of school management and pedagogic practices in a particular school. The SSME was designed to capture indicators that are believed to affect student learning. The results of SSME surveys can inform school, district, provincial, or national administrators and donors about current practices in their schools and classrooms.

SSME collects information on (1) basic school inputs such as infrastructure, learning materials, teacher and school director characteristics, student characteristics, and parental and community involvement; and (2) classroom teaching practices, including use of material, instructional content, student teacher interaction, assessment methods, and administrative oversight. When analyzed together, these instruments create a comprehensive picture of a school's learning environment.

CHAPTER 2: METHODOLOGY

This chapter describes the RFS-SSE sampling process, development of the EGRA assessment tool, preparation and execution of data collection, and analysis of the quality of the tool.

2.1 RFS-SSE Sampling Design

A stratified cluster random sampling method was used in order to assure that (1) an equal number of boys and girls in urban and rural schools would be assessed and (2) that the results of the study would be generalizable to the entire population of intervention schools in each of the eight delegations selected for intervention. Schools were first stratified by geographic location and urban/rural environment. Within schools, students were stratified by gender. The detailed steps of this process are described below.

The project is piloting in a total of 90 schools throughout eight delegations in Morocco. In each of the eight RFS-SSE delegations, five to seven intervention schools were selected to be a part of the Baseline study sample. The following is a description of the process used for the selection of schools and students. A database of the schools in each delegation served as the databank from which the school sample was pulled. The Ministry of Education's "Massar" (Education Management Information Systems or EMIS) database was used for the student sample.

1. All 1415 schools in each of the eight delegations were stratified based on location (rural/urban).
2. Schools were deselected if they contained multi-level Grade 1 or Grade 2 classrooms.
3. Between 10 and 12 schools were selected per delegation to form the pool of 90 intervention schools for the RFS-SSE project.
4. Among the 90 intervention schools, five to seven schools were randomly selected in each region based on location to form the sample of treatment schools for the EGRA and SSME study. Therefore, half (i.e., 45) of the intervention schools are included in the baseline study.
5. For the EGRA and SSME study, an additional five to seven schools per delegation were selected to form the control group. Therefore, another 45 schools make up the control group.
6. Within each school, a sample of 10 Grade 1 girls and 10 Grade 1 boys was selected for a total of 20 Grade 1 students per school. In schools where fewer than 20 Grade 1 students were enrolled, all Grade 1 students present on that day were assessed, resulting in a boy-girl ratio of 52:48. In schools with more than 20 enrolled Grade 1 students, a list of randomly selected replacement students in each school was also provided by the Ministry of Education.

To enhance comparability over time, the same schools will be included at all four evaluation time points. Therefore, the Midline 1 was conducted in the same schools as the Baseline. Additionally, because the study is longitudinal, the same students were assessed within each school at both time points (Baseline and Midline 1).

2.2 School and Student Sample

School sample

At the time of Baseline data collection, schools within each delegation were randomly selected with equal probability from the MOE's EMIS database. The sample of schools that participated in the Baseline and Midline 1 studies came from four AREFs (Académie Régionale d'Education et de Formation) and from two delegations per AREF, making a total of eight delegations. Within each selected school, students were randomly selected using the MOE's EMIS system. The final Baseline sample included 1,729 students in 90 schools (45 control and 45 experimental). The same 90 schools were included in Midline 1.

Student sample

At the time of Midline 1 data collection, the same schools and students included at Baseline were assessed again. However, a small number of students who participated in the Baseline study were not available, reducing the final Midline 1 sample to only 1,677 students (97% of the Baseline sample). The proportion of students lost between Baseline and Midline 1 was a little larger in the control group than in the experimental group. In the control group, the follow-up rate was 96% (840 reached out of 871 target) while it was 98% (837 reached out of 858 target) in the experimental group. The difference between the two groups was minimal and does not create any bias in the results. Each delegation was able to retain a similar proportion of students in the control group as in the experimental group. In total, in all the delegations, 50.1 % of the students were a part of the control group and 49.9% were a part of the experimental group (see Table 1).

TABLE 1. NUMBER OF STUDENTS BY AREF AND BY DELEGATION

AREF / Region	**Delegation / Province**	**Control Group** n (%)	**Experimental Group** n (%)	**Total** n (%)
Sous Massa	Inezgane-Aït Melloul	117 (50.00%)	117 (50.00%)	234 (100%)
	Tiznit	109 (47.19%)	122 (52.81%)	231 (100%)
Rabat-Kénitra	Témara	99 (50.51%)	97 (49.49%)	196 (100%)
	Kenitra	96 (50.26%)	95 (49.74%)	191 (100%)
Orient	Oujda-Angad	115 (49.57%)	117 (50.43%)	232 (100%)
	Figuig	115 (52.51%)	104 (47.49%)	219 (100%)
Fès Meknès	Taounate	97 (49.49%)	99 (50.51%)	196 (100%)
	Elhajeb	92 (51.69%)	86 (48.31%)	178 (100%)
Total		**840 (50.09%)**	**837 (49.91%)**	**1,677 (100%)**

At Midline 1, the proportion of female and male students remained similar to the Baseline. Ninety-seven percent of male students from Baseline and 97 percent of female students from Baseline were reached at Midline 1. Table 2 provides a breakdown by sex and study group of the students who were reached at Midline 1.

TABLE 2. NUMBER OF STUDENTS BY SEX AND BY GROUP

	Girls n (%)	**Boys** n (%)	**Total** n (%)
Control Group	445 (52.98%)	395 (47.02%)	840 (100%)
Experimental Group	431 (51.49%)	406 (48.51%)	837 (100%)
Total	**876 (52.24%)**	**801 (47.76%)**	**1,677 (100%)**

Finally, as for distribution according to the location of the school, this also remained the same as at Baseline. The follow-up rate in both rural and urban areas was 97%. Table 3 shows the breakdown by school location of the students who were reached at Midline 1.

TABLE 3. NUMBER OF STUDENTS BY SCHOOL LOCATION AND GROUP

	Rural n (%)	**Urban** n (%)	**Total** n (%)
Control Group	413 (49.17%)	427 (50.83%)	840 (100%)
Experimental Group	410 (48.98%)	427 (51.02%)	837 (100%)
Total	**823 (49.08%)**	**854 (50.92%)**	**1,677 (100%)**

Therefore, the characteristics of the Midline 1 sample group suggest that there was no differential loss of subjects since Baseline as a function of group, sex, or location of the school. We can conclude that the follow-up rate for Midline 1 is very high and there should be no bias caused by a differential loss of subjects.

2.3 EGRA/SSME Tool Development

Prior to Baseline data collection, an in-depth tool development process was completed in order to develop the EGRA assessments for Baseline and Midline 1 and the SSME surveys.[1] In order to ensure comparability across time points, the tools administered at Midline 1 were based on the Baseline tools. Slight modifications were made to the Baseline SSME survey instruments as appropriate prior to administering them again at Midline 1.

The Midline 1 EGRA consisted of the same six subtasks as the Baseline. The first three subtasks (phonemic awareness, syllable identification, and nonword reading) contained the same items as the Baseline assessment, but the order of the items was rearranged within each subtask. The stories and related comprehension questions for the last three subtasks (passage reading, reading comprehension, and listening comprehension) were different than those used at Baseline. The stories used in the Baseline and Midline 1 tools were all developed at the same time, during the original tool development process prior to the Baseline. At that time, piloting of the stories was conducted, and two of the piloted stories were chosen for the reading comprehension subtask and the listening comprehension subtask. One of the stories for each subtask was used in the Baseline tool and the other story for each subtask was used in the Midline 1 tool.

The EGRA tool was developed to measure six core reading competencies in Grades 1 and 2. Some of the subtasks were timed in order to measure the speed with which students were able to correctly identify graphemes (syllable identification), decode nonwords (phonics), and read connected text (fluency). Below in Table 4 is a summary of the EGRA subtasks and the skills they are designed to assess.

TABLE 4. EGRA SUBTASKS AND SKILLS[2]

EGRA Subtask	Core Reading Skill	Untimed/ Timed	Skill Demonstrated by Students' Ability To:
1. Phonemic Awareness	Phonological Awareness	Untimed	Identify the initial sounds of words by having the assessor and then the student read the phonemes aloud
2. Syllable Identification	Alphabet Knowledge	Timed (1 Minute)	Provide the sound of syllables presented in a random order
3. Nonword Reading	Decoding	Timed (1 Minute)	Make letter-sound correspondences (grapheme -phoneme correspondences) through the reading of simple nonsense words
4a. Passage Reading	Oral Reading Fluency	Timed (1 Minute)	Read a text with accuracy, with little effort, and at a sufficient rate

[1] A detailed explanation of the tool development process is presented in "USAID/Morocco: Reading for Success – Small-Scale Experimentation Activity (RFS-SSE) Early Grade Reading Assessment Baseline Report" prepared by School-to-School International and Chemonics International, Inc. in May 2016.

[2] This table is taken from the EGRA Toolkit 2.0 (RTI International. 2015. *Early Grade Reading Assessment (EGRA) Toolkit, Second Edition*. Washington, DC: United States Agency for International Development.)

4b. Reading Comprehension	Reading Comprehension	Untimed	Respond correctly to different types of questions, including literal and inferential questions about the text they have read
5. Listening Comprehension	Listening Comprehension; Oral Language	Untimed	Respond correctly to different types of questions, including literal and inferential questions about the text the assessor reads to

Phonemic Awareness: On this untimed subtask, students were asked to orally identify a phoneme (the smallest unit of sound in a word) at the beginning of 10 familiar words.

Syllable Identification: On this timed subtask, students were presented with 100 letter-diacritic combinations and asked to read the syllables in one minute.

Nonword Reading: On this timed subtask, students were presented with 50 nonwords (words invented for this exercise) and asked to read as many of the words as possible using decoding skills in one minute.

Passage Reading: On this timed subtask, students were given one minute to read a passage consisting of 61 words.

Reading Comprehension: On this untimed subtask, students were asked comprehension questions based on how far into the 61 word passage they were able to read in the Passage Reading subtask. Students were only asked questions for which they had read far enough in the passage to come across the answer.

Listening Comprehension: On this untimed subtask, students were first read a story aloud by the enumerators, then asked five questions orally to assess their understanding of the story's meaning.

2.4 Data Collection

Prior to Midline 1 data collection, RFS-SSE conducted a three-day refresher training for enumerators from May 9-11, 2016 in Rabat. All Midline 1 enumerators had also served as enumerators during the Baseline study and had successfully completed a seven-day enumerator training in January 2016. The enumerators were selected by the MOE and were equally sourced from each of the eight delegations of the RFS-SSE activity. During the refresher training in May, the enumerators were re-trained to administer all six EGRA subtasks on tablets using the electronic data capture application Tangerine. Those who acted as supervisors during Baseline data collection were also re-trained to administer all six SSME tools using Tangerine and to supervise data collection teams.[3]

[3] One of the Baseline supervisors was unavailable for participation in the midline 1 data collection. The replacement supervisor received a day-long one-on-one training on SSME administration and data collection team supervision including a visit to a nearby school to practice administration of the SSME tools.

Enumerators were re-trained on the specific contents of each of the EGRA subtasks, proper administration protocols for each subtask, and the use of tablets and the Tangerine application, among other topics. Enumerator performance was monitored regularly throughout training by the two lead facilitators. Facilitators led two sessions to measure consistency of scoring across all enumerators. Enumerator consistency during training ranged from 96-100% agreement depending on the subtask (see Annex B).

Eight new provincial representatives were also trained as Provincial Field Coordinators (PFCs) during the refresher training and during an additional half-day training conducted exclusively for them.

During training, the enumerators and PFCs conducted practice in neighboring schools. In the practice session, teams of enumerators practiced administering the EGRA tools with Grade 1 students. The eight PFCs, the training facilitators, and key members of the RFS-SSE team oversaw the enumerators' performance. During this school visit, supervisors also practiced administering the SSME teacher observation tool and performing their duties as team leaders.

Data collection

Data collection occurred from May 16-27, 2016 and the enumerator teams were able to reach approximately 97% of targeted students throughout the 90 schools. To ensure the quality of data collection, a trained supervisor oversaw each data collection team of two EGRA enumerators. At the end of each day, supervisors and enumerators discussed progress and problems encountered that day. The supervisors verified that each enumerator had reached as many of the Baseline participants as possible at each school and ensured that a daily data collection report was completed. Each team visited one to two schools per day. Because of the longitudinal study design, it was occasionally necessary to re-visit schools to assess students who were absent during the first scheduled school visit. When possible, schools with students missing during the first day of data collection were visited a second time in order to try to reach the missing students. If the students were also missing on the second school visit, then the students were marked as "absent" in our dataset and not included in the Midline 1 sample. The teams uploaded the data from their tablets to a cloud database each evening of data collection. Those data were reviewed and tallied nightly by the STS EGRA Coordinator and discussed with the PFCs as necessary.

Inter-rater reliability during data collection

In order to record the level of enumerator agreement throughout data collection, assessors undertook daily measures of inter-rater reliability (IRR) according to the following protocol. Enumerators

worked in pairs to assess the first student of the day. During this first assessment, one enumerator acted as the "main enumerator", administering the EGRA as usual, and scoring the student responses in his or her tablet. The second enumerator simultaneously listened and also scored the student responses in his or her tablet. Once the assessment was completed and the student had returned to class, the two enumerators compared and discussed their scoring of the student's responses. Any points of disagreement or difference in marking were brought to the attention of supervisors and discussed during team meetings. Enumerator pairs took turns playing the role of main enumerator from each day to the next. The percent agreement between enumerators throughout Midline 1 data collection is presented below by subtask in Table 5. This is a measure of how many items the enumerators within a pair scored differently on an assessment out of the total number of items scored. Additionally, Kappa statistics are also reported in Table 5. Kappa measures the extent to which two different scorings of the same student could have happened by chance. Kappa values greater than .75 are considered excellent. Overall, among the 88 instances of IRR administration conducted during data collection, the agreement was very high. Therefore, we can be confident that enumerators were scoring consistently throughout data collection.

TABLE 5. INTER-RATER RELIABILITY DURING DATA COLLECTION

Subtask Name	Percent Agreement	Kappa
Phonemic Awareness	99.31%	0.99
Syllable Identification	99.44%	0.97
Nonword Reading	99.45%	0.95
Passage Reading	99.52%	0.97
Reading Comprehension	99.32%	0.93
Listening Comprehension	99.55%	0.99

2.5 Quality of Assessment Tool

Analyzing the qualities of the assessment ensures that the conclusions drawn about the performance of students on the assessment are valid. The first analysis of the quality of the assessment consisted of verifying the accuracy, or the reproducibility, of scores coming from each of the subtasks. The Cronbach's alpha coefficient is used to verify the internal consistency of each of the subtasks. This index varies between 0 and 1, where a value closer to 1 indicates that the performance of the students can be easily generalized. In other words, each subtask represents just a sample of items among the universe of total possible items. For example, the subtask of phonemic awareness is made up of 10 items, but the universe of total possible phonemes is larger. A value of Cronbach's alpha coefficient close to 1 allows us to be confident that the performance of a student would have been similar even with a sample of 10 different phonemes. Cronbach's alpha coefficients for the Midline 1 EGRA tool are present below in
Table 6.

Just as was observed at Baseline, for several subtasks (phonemic awareness, syllable identification and nonword reading), the value of the Cronbach's alpha coefficient is very high. The two comprehension subtasks (reading and listening) had lower values. The limited number of items in each of these subtasks could explain the lower values. On reading comprehension, the low number of items (5) and – based on the structure of the task – the number of students who answered all questions yields a low value for Cronbach's alpha. The values observed for the Midline 1 assessment were similar to those at Baseline. Cronbach's alpha cannot be calculated for the passage reading subtask because the story, in its entirety, must be considered as one item. Because the reading passage is one unique item, no generalization can be made to a universe of similar items.

TABLE 6. INDEX OF INTERNAL CONSISTENCY BY TASK AND FOR THE ENTIRE EGRA

Tasks	Number of Items	Cronbach's Alpha
Phonemic Awareness	10	0.968
Syllable Identification	100	0.985
Nonword Reading	50	0.951
Reading Comprehension	5	0.489
Listening Comprehension	5	0.699

For each subtask, an item analysis can also identify problematic items that should be excluded from the total score on each subtask. These analyses are based on two indices: the difficulty of the item (p) and its discrimination (d). The difficulty index represents the proportion of students who gave a correct answer. Ideally, this index should not have a value that is too low (item is too difficult) or too high (item is too easy). As for the discrimination index, it represents the correlation between the score on an item and the total score on that subtask. For this index, the values should not be negative or too close to 0. The results of this analysis by subtask are presented in ANNEX A: Reliability of the EGRA Subtask Items

Based on this analysis, no items appeared to be problematic. It should be noted that for the timed subtasks (syllable identification and nonword reading), the last items presented values that were low since few of the students had the time in one minute to get to these items. The same phenomenon was observed for the reading comprehension subtask. Given that the comprehension questions were linked to the amount of the passage read, the last comprehension questions were attempted by only a very small number of students (the majority were not able to finish reading the passage during the minute that the task was allotted and, therefore, were not asked the final comprehension questions).

2.6 Data Analysis

Like the Baseline report, the following analysis steps were conducted: 1) reliability estimates of each subtask, 2) subtask and item statistics, and 3) timed and untimed subtask scores. Since the same children were again assessed at Midline 1, paired comparisons of subtask scores between Baseline and Midline 1 were computed using a difference score for each subtask. Mean scores for the Baseline EGRA have been updated throughout this report to include only those students for whom Midline 1 data was also captured. This allows for a direct comparison of the change in scores for the same exact students at both time points. Comparisons of the same students between mean difference scores for control and experimental groups is how we measure the impact of the treatment on the students. Data for the SSME questionnaires were correlated with the difference scores of each subtask to identify schools or students' characteristics that are linked to the change in subtask scores between Baseline and Midline 1.

CHAPTER 3: EGRA RESULTS

Detailed EGRA results of the Midline 1 are presented below, including the average score by subtask for all students as well as the average score disaggregated by study group (experimental and control), the proportion of students unable to answer a single item correctly by subtask, and a comparison of schools in rural and urban areas. The difference in mean score between the Baseline and Midline 1 is also reported. It is important to keep in mind that the difference in mean score between Baseline and Midline 1 includes only the students who participated in both evaluations. This gives us a true measure of the change in students' scores that can be attributed to the intervention.

The RFS-SSE activity has several key indicators directly related to the EGRA subtasks included on the assessment for Grade 1 students. Table 7 and

Fluency measures	1. Average Number of Correct Items Per Minute (Fluency)				2. Percent of Students Receiving Zero Score			
	Baseline		**Midline 1**		**Baseline[5]**		**Midline 1**	
	Experimental	Control	Experimental	Control	Experimental	Control	Experimental	Control
Syllable Identification	23.27	23.13	37.60	37.09	15.89%	16.43%	8.48%	8.33%
Nonword Reading	7.20	6.86	11.41	11.52	39.78%	41.31%	24.49%	28.10%
Passage Reading	5.83	5.59	13.28	13.04	45.28%	45.71%	21.62%	25.12%

Table 8 below report the Baseline and Midline 1 data of the experimental schools only for the relevant indicators for Intermediate Result 1, "Effectiveness of Arabic reading lessons on student reading competencies tested for Grades 1 and 2."[4]

TABLE 7. BASELINE AND MIDLINE 1 DATA FOR THE RELEVANT INDICATORS FOR INTERMEDIATE RESULT 1 – FLUENCY MEASURES

Fluency measures	1. Average Number of Correct Items Per Minute (Fluency)				2. Percent of Students Receiving Zero Score			
	Baseline[5]		**Midline 1**		**Baseline[5]**		**Midline 1**	
	Experimental	Control	Experimental	Control	Experimental	Control	Experimental	Control
Syllable Identification	23.27	23.13	37.60	37.09	15.89%	16.43%	8.48%	8.33%
Nonword Reading	7.20	6.86	11.41	11.52	39.78%	41.31%	24.49%	28.10%
Passage Reading	5.83	5.59	13.28	13.04	45.28%	45.71%	21.62%	25.12%

TABLE 8. BASELINE AND MIDLINE 1 DATA FOR THE RELEVANT INDICATORS FOR INTERMEDIATE RESULT 1 – COMPREHENSION MEASURES

Comprehension measures	3. Average Correct Answers To Comprehension Questions				4. Percent of Students Receiving Zero Score			
	Baseline[5]		**Midline 1**		**Baseline[5]**		**Midline 1**	
	Experimental	Control	Experimental	Control	Experimental	Control	Experimental	Control

[4] Additional details on the RFS-SSE indicators can be found in the Monitoring, Evaluation and Learning Plan updated in March 2016.

[5] Mean scores of only the students who were assessed at both Baseline and Midline 1.

Reading Comprehension	.36	.38	.31	.36	75.39%	76.67%	77.42%	76.67%
Listening Comprehension	1.13	1.15	1.75	1.59	45.64%	46.43%	28.32%	31.31%

3.1 EGRA Results by Subtask

Phonemic awareness

The phonemic awareness subtask is made up of 10 items. Since this subtask was not timed, the students had the opportunity to provide a response to all 10 items unless a student was not able to correctly answer any of the first five items. In that case, the subtask was stopped early and the student moved on to the next subtask. At Midline 1, the control group, on average, correctly identified 4.89 phonemes out of 10, compared to the experimental group which, on average, correctly identified 6.19 phonemes out of 10. Comparing the results of Baseline to Midline 1, we observed that the students from the control group were, on average, able to recognize 1.56 more phonemes at Midline 1 compared to Baseline, while the students from the experimental group were, on average, able to recognize 3.08 more phonemes (see Table 9). The difference between the progress of the two groups is statistically significant ($t(1675)=-5.33$; $p<0.000$). The effect size of this difference is 0.35 or a difference of little to medium size between the two groups.

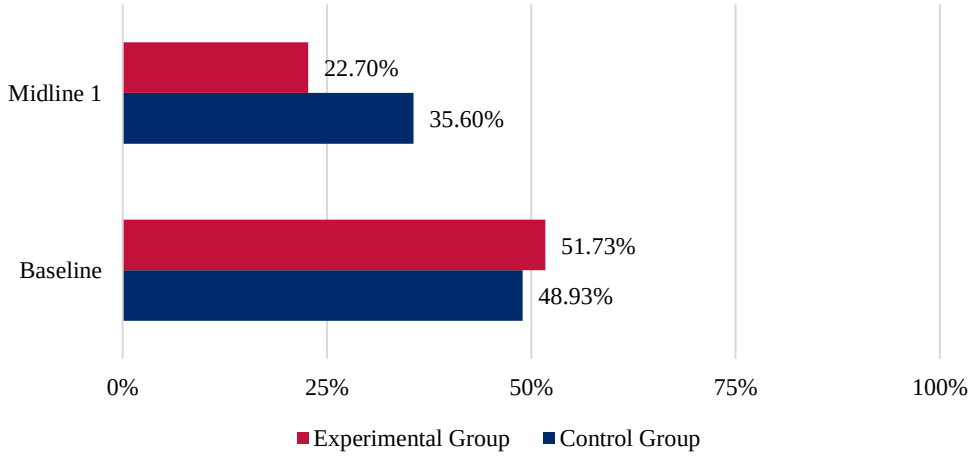
TABLE 9. CORRECT RESPONSES ON THE PHONEMIC AWARENESS SUBTASK

	Results Obtained at Midline 1				Difference from Baseline	
	n[6]	**Mean[7]**	**Standard Deviation[8]**	**Range[9]**	**Mean**	**Standard Deviation**
Control Group	840	4.89	4.42	0 – 10	1.56	4.29
Experimental Group	837	6.19	4.16	0 – 10	3.08	4.45
Total	**1677**	**5.54**	**4.34**	**0 – 10**	**2.32**	**4.43**

The proportion of students receiving zero scores on the phonemic awareness subtask decreased between Baseline and Midline 1. For the control group, 35.60% (n=299) of the students had no correct answers on this subtask at Midline 1, while for the experimental group, 22.70% (n=190) of the students had no correct answers (see Figure 1). The difference between the two groups is statistically significant ($\chi^2(1,677) = 33.75$, $p<0.01$).

FIGURE 1. PROPORTION OF STUDENTS RECEICING ZERO SCORES ON THE PHONEMIC AWARENESS SUBTASK

[6] n="number" (size of sample).
[7] Mean=average score.
[8] Standard deviation=measure of the variation in a distribution.
[9] Range=the difference between the largest and smallest values.



Syllable identification

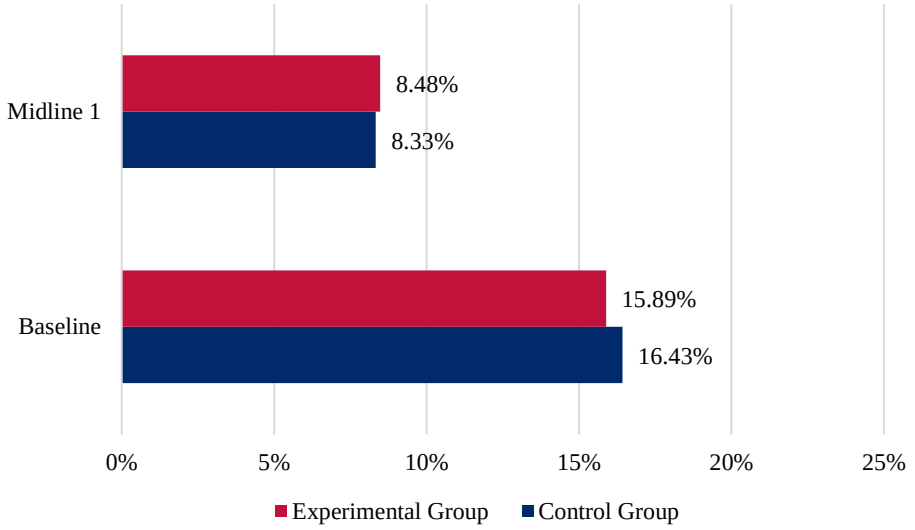
Syllable identification was the second subtask of the EGRA. This subtask was timed and included a total of 100 syllables. If a student did not correctly answer any of the first 10 items, then the subtask was stopped and the student was moved on to the next subtask. At Midline 1, students from the control group were able to recognize, on average, 37.09 syllables per minute while students from the experimental group were able to recognize, on average, 37.60 syllables per minute. The progress between Baseline and Midline 1 was about 14 more syllables per minute for both groups (see Table 10). The difference in progress between the two groups is not statistically significant (t(1)=0.30; p=0.762).

TABLE 10. CORRECT RESPONSES ON THE SYLLABLE IDENTIFICATION SUBTASK

	Results Obtained at Midline 1				Difference from Baseline	
	n	**Mean**	**Standard Deviation**	**Range**	**Mean**	**Standard Deviation**
Control Group	840	37.09	26.12	0 – 99	13.96	13.98
Experimental Group	837	37.60	24.76	0 - 108	14.33	14.15
Total	**1677**	**37.34**	**25.44**	**0 – 108**	**14.14**	**14.06**

The proportion of students receiving zero scores on the syllable identification subtask decreased from Baseline to Midline 1. At Midline 1, 8.33% (n=70) of students from the control group were unable to correctly identify at least one syllable. This proportion was similar for the experimental group in which 8.48% (n=71) of students were unable to correctly identify at least one syllable (See Figure 2). The difference between the two groups is not statistically significant. (χ^2 (1,1677)=0.012, p=0.912).

FIGURE 2. PROPORTION OF STUDENTS RECEIVING ZERO SCORES ON THE SYLLABLE IDENTIFICATION SUBTASK



Nonword reading

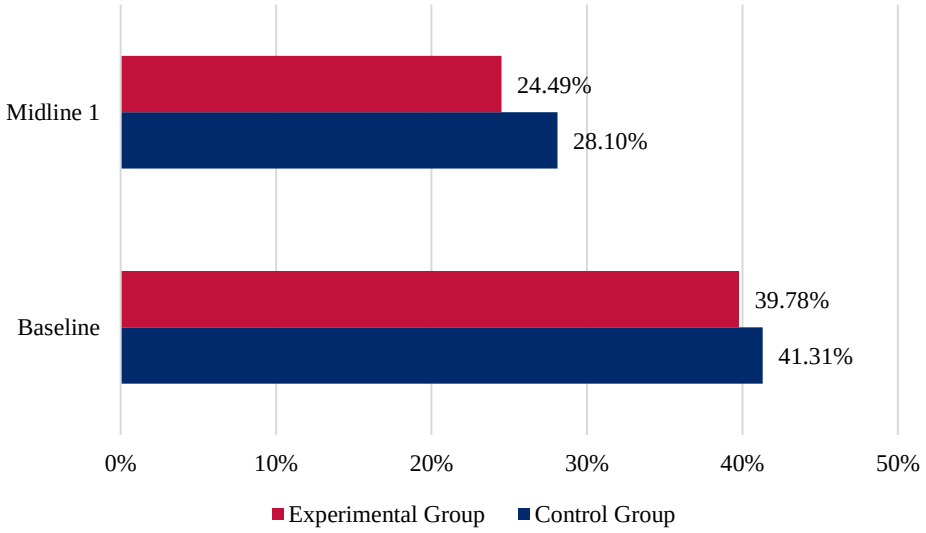
Similar to the syllable identification subtask, the nonword reading subtask was also timed and students were given up to one minute. If a student was unable to correctly read any of the first five nonwords, then the subtask was stopped and the student was moved on to the next subtask. This subtask included a total of 50 nonwords. For this subtask, the student could have read up to 50 total nonwords (or invented words) in one minute. The greatest number of nonwords actually read in one minute by a student was 40 words. The students from both the control group and the experimental group read, on average, about 11 words per minute. There was no statistically significant difference between the averages of the two groups and both groups' progress between the Baseline and Midline 1 was similar. On average, students from the control group read 4.66 more words in one minute at Midline 1 while students from the experimental group read 4.21 more words in one minute (See Table 11). The difference in progress between the two groups is not statistically significant ($t(1675)=0.936$, $p=0.352$).

TABLE 11. CORRECT RESPONSES ON THE NONWORD READING SUBTASK

	Results Obtained at Midline 1				Difference from Baseline	
	n	**Mean**	**Standard deviation**	**Range**	**Mean**	**Standard deviation**
Control Group	840	11.52	9.86	0 – 37	4.66	5.65
Experimental Group	837	11.41	9.51	0 – 40	4.21	5.66
Total	**1677**	**11.47**	**9.68**	**0 – 40**	**4.44**	**5.66**

The proportion of students receiving zero scores on the nonword reading subtask decreased from Baseline to Midline 1. In the control group, 28.10% (n=236) of students were not able to read at least one nonword while in the experimental group, 24.49% (n=205) of students were not able to read at least one nonword (see Figure 3). The difference between the two groups is not statistically significant. (χ^2 (2, 1677)=2.81, p=0.094).

FIGURE 3. PROPORTION OF STUDENTS RECEIVING ZERO SCORES ON THE NONWORD READING SUBTASK



Oral reading passage

The oral reading passage for the Midline 1 consisted of 61 words total, and students were given up to one minute to read the entire passage. If a student was unable to correctly read any of the first five words, then the subtask was stopped and the child was moved on to the next subtask.

An oral reading fluency (ORF) rate is calculated for this subtask by taking into consideration the number of words read aloud by the student, how many of those words were read correctly, and how much time it took them. The number of Correct Words Per Minute (CWPM) is calculated using this information.

At Midline 1, only one student was able to read the entire passage in less than one minute.[10] When a student is able to read the entire passage in *less than* one minute, the Correct Words Per Minute score is adjusted to calculate how many total words the student *could* have read in a full minute if there were more words in the passage. This calculation assumes that the student would have read at the same pace throughout the entire minute if they had a longer text. Because this was the only student who read the passage in less than a minute, he had the highest oral reading fluency rate (62.1 words per minute) out of all students assessed at Midline 1.

The students from the control group were able to read, on average, 13.04 CWPM while those from the experimental group read, on average, 13.28 CWPM. The difference between the two groups was not statistically significant. On average, both groups were able to read 7.45 more words correctly per minute at Midline 1 compared to Baseline. Thus, the progress of the students from the control and experimental groups was not different (see
Table 12).

TABLE 12. CORRECT RESPONSES ON THE PASSAGE READING SUBTASK

	Results Obtained at Midline 1	Difference from Baseline

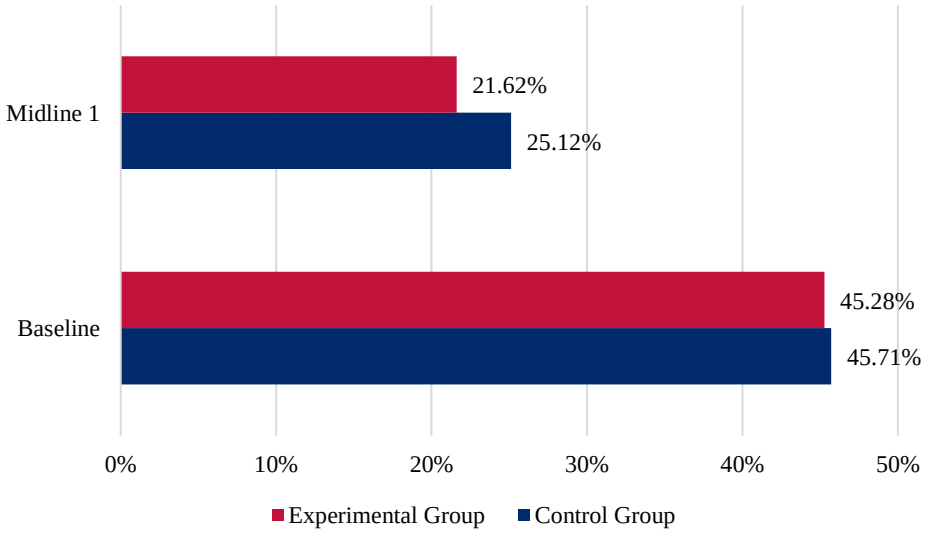
[10] This was not the same student who read the entire passage in less than one minute at Baseline.

	n	Mean	Standard deviation	Range	Mean	Standard deviation
Control Group	840	13.04	12.22	0 – 62.1	7.45	7.18
Experimental Group	837	13.28	12.04	0 – 58	7.45	7.19
Total	**1677**	**13.16**	**12.12**	**0 – 62.1**	**7.45**	**7.19**

The number of students receiving a zero score on the passage reading subtask decreased since Baseline. For the control group, 25.12% (n=211) of the students were unable to correctly read at least one word, and this proportion was 21.62% (n=181) for the experimental group (see

Figure 4). The difference between the two groups is not statistically significant (χ^2 (1,677)=2.86, p=0.91).

FIGURE 4. PROPORTION OF STUDENTS RECEIVING ZERO SCORES ON THE PASSAGE READING SUBTASK



Reading comprehension

The reading comprehension subtask contains a total of five comprehension questions about the reading passage read aloud by the student in the previous subtask. Students were asked only the comprehension questions for which they had read far enough in the passage to be able to know the answer. Therefore, students were not always asked all five of the comprehension questions. For

example, the answer to comprehension question #4 was not revealed in the reading passage until the fifth sentence of that passage. Therefore, if a student only read through the first four sentences, then he/she would only be asked the first three questions. The students who were not able to read a single word of the reading passage (around 23% of the total sample as noted in

Figure 4) were not asked any of these questions and the student received a zero task on this subtask.

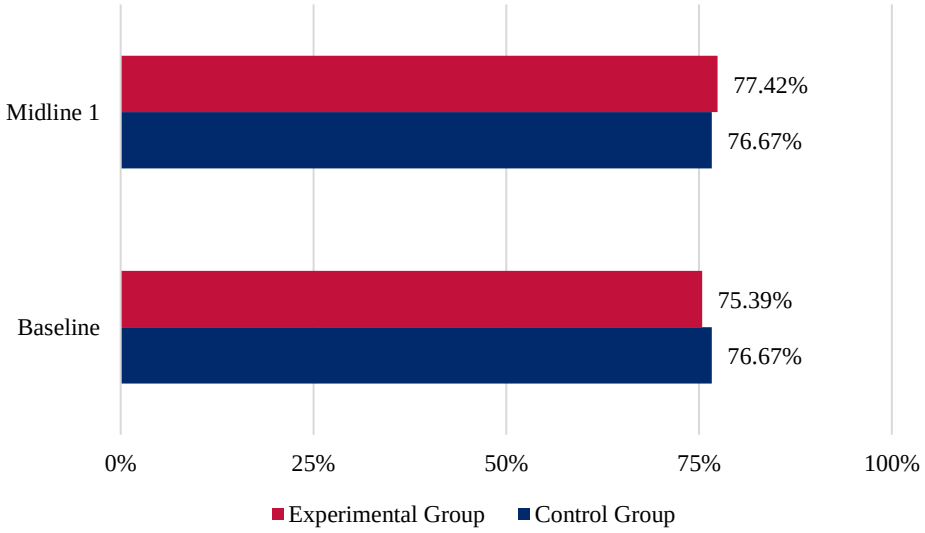
Students from both the control and experimental groups were asked, on average, the same number of questions (meaning that all students were able to read approximately the same amount of the passage). The students from the control group were able to answer, on average, 0.36 questions correctly while students from the experimental group were able to answer, on average, 0.31 questions correctly. Of the five questions asked, the students were able to give four correct answers. The progress since Baseline of the students from both groups is not statistically significant ($t(1675)=0.75$, $p=0.452$) (see Table 13).

TABLE 13. CORRECT ANSWERS ON THE READING COMPREHENSION TASK

	Results Obtained at Midline 1				Difference from Baseline	
	n	**Mean**	**Standard Deviation**	**Range**	**Mean**	**Standard Deviation**
Control Group	840	0.36	0.73	0 – 4	-0.02	0.65
Experimental Group	837	0.31	0.65	0 – 3	-0.05	0.58
Total	**1677**	**0.34**	**0.69**	**0 – 4**	**-0.04**	**0.61**

The proportion of students not able to answer at least one question correctly remained the same as Baseline for both the control and experimental groups (see Figure 5). While the proportion of students receiving zero scores appears higher in the experimental group at Midline 1 (77.42%, n=648) compared to Baseline (75.39%, n=631), the results are not statistically significant and, therefore, cannot be interpreted as a true change. The difference between the two groups was not statistically significant ($\chi^2(1,1677)=0.13$, $p=0.714$).

FIGURE 5. PROPORTION OF STUDENTS RECEIVING ZERO SCORES ON THE READING COMPREHENSION SUBTASK



Listening comprehension

Listening comprehension was the last subtask of the EGRA. For this subtask, students were read a short story aloud by the enumerator and then were asked five comprehension questions about the story. All students were asked all five questions (there is no "auto-stop" feature) on this subtask. On average, students from the control group were able to correctly answer 1.59 questions out of five. For the experimental group, on average, students were able to correctly answer 1.75 questions out of five. The number of correct answers increased since Baseline, on average, more in the experimental group than in the control group. Students from the control group provided, on average, 0.44 more correct answers at Midline 1 compared to Baseline, while those from the experimental group provided, on average, 0.62 more correct answers (see Table 14). The difference between the two groups is statistically significant ($t(1675)=2.36$, $p=0.021$). The effect size of this difference is 0.14. This figure indicates that the difference between the two groups was small.

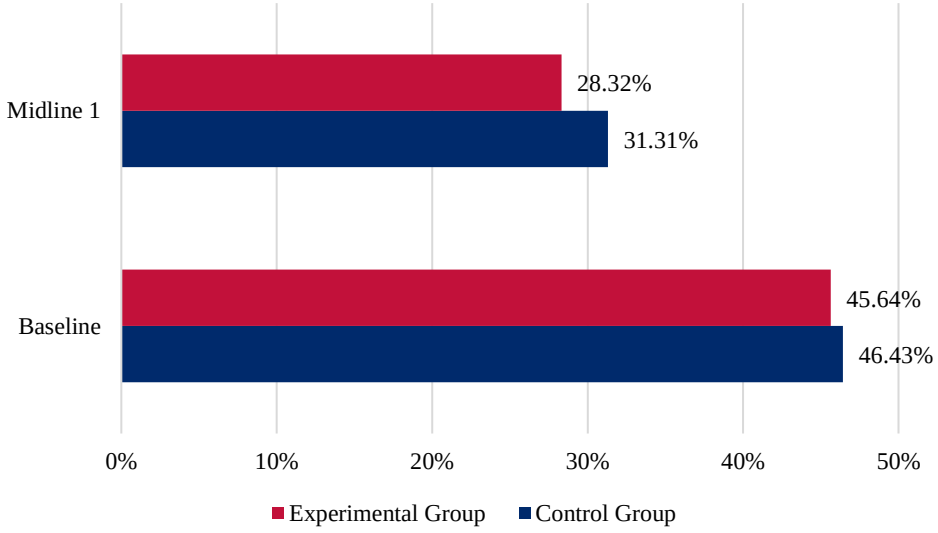
TABLE 14. CORRECT ANSWERS ON THE READING COMPREHENSION SUBTASK

	Results Obtained at Midline 1				Difference from Baseline	
	n	**Mean**	**Standard Deviation**	**Range**	**Mean**	**Standard Deviation**
Control Group	840	1.59	1.50	0 – 5	0.44	1.29
Experimental Group	837	1.75	1.56	0 – 5	0.62	1.30
Total	**1677**	**1.67**	**1.53**	**0 – 5**	**0.53**	**1.30**

The proportion of students receiving zero scores decreased in both the control and experimental groups compared to Baseline. For the control group, 31.31% (n=263) of students were unable to provide a single correct response; for the experimental group, 28.32% (n=237) of students were unable to provide a single correct response (see Figure 6). The difference between the two groups is not statistically significant (χ^2 (1,1677)=0.11, p=0.746).

FIGURE 6. PROPORTION OF STUDENTS RECEIVING ZERO SCORES ON THE LISTENING COMPREHENSION

	Total				Rural				Urban				p
	Control Group		**Exp. Group**		**Control Group**		**Exp. Group**		**Control Group**		**Exp. Group**		
	Mean	**SD**	**Mean**	**SD**	**Mean**	**SD**	**Mean**	**SD**	**Mean**	**SD**	**Mean**	**SD**	
Phonemic Awareness	1.56	4.29	3.08	4.45	1.49	4.13	3.39	4.35	1.64	4.44	2.78	4.53	0.406
Syllable Identification	13.96	13.98	14.33	14.15	14.18	14.06	15.21	14.47	13.74	13.91	13.48	13.08	0.303
Nonword Reading	4.66	5.65	4.21	5.66	4.88	5.91	4.48	5.63	4.44	5.39	3.95	5.68	0.239
Passage Reading*	7.45	7.18	7.45	7.19	6.13	7.17	6.00	6.84	8.73	6.97	8.83	7.25	<0.00
Reading Comprehension*	0.02	0.65	-0.05	0.58	0.04	0.48	0.00	0.39	-0.08	0.77	-0.11	0.71	0.004
Listening Comprehension	0.44	1.29	0.62	1.30	0.43	1.18	0.59	1.16	0.44	1.39	0.64	1.42	0.716



3.2 EGRA Results by Urbanicity

Type of school location (Urban/rural)

At Baseline, the students living in urban areas obtained higher EGRA results compared to those living in rural areas. This was true for all of the EGRA subtasks. This difference was more significant for the syllable identification, nonword reading, and passage reading subtasks. The following table shows the progress of the students between Baseline and Midline 1 by school location.

Generally, the progress of students since Baseline was similar regardless of whether they attended school in a rural or urban setting. On the other hand, the passage reading and reading comprehension subtasks did show a statistically significant difference in progress as a function of the school location. For the passage reading subtask, students in urban schools were able to read an average of 2.6 additional words in one minute compared to students in rural schools. This difference was similar whether the student was part of the control group or the experimental group (see Table 15). There was no difference found in the performance of students in urban or rural areas for the remaining subtasks.

TABLE 15. DIFFERENCE FROM BASELINE BY SUBTASK AND SCHOOL LOCATION

	Total				Rural				Urban				p
	Control Group		**Exp. Group**		**Control Group**		**Exp. Group**		**Control Group**		**Exp. Group**		
	Mean	**SD**	**Mean**	**SD**	**Mean**	**SD**	**Mean**	**SD**	**Mean**	**SD**	**Mean**	**SD**	
Phonemic Awareness	1.56	4.29	3.08	4.45	1.49	4.13	3.39	4.35	1.64	4.44	2.78	4.53	0.406
Syllable Identification	13.96	13.98	14.33	14.15	14.18	14.06	15.21	14.47	13.74	13.91	13.48	13.08	0.303
Nonword Reading	4.66	5.65	4.21	5.66	4.88	5.91	4.48	5.63	4.44	5.39	3.95	5.68	0.239
Passage Reading*	7.45	7.18	7.45	7.19	6.13	7.17	6.00	6.84	8.73	6.97	8.83	7.25	<0.00
Reading Comprehension*	0.02	0.65	-0.05	0.58	0.04	0.48	0.00	0.39	-0.08	0.77	-0.11	0.71	0.004
Listening Comprehension	0.44	1.29	0.62	1.30	0.43	1.18	0.59	1.16	0.44	1.39	0.64	1.42	0.716

*Statistically significant difference ($p<0.05$)

Relationships between the results of the EGRA subtasks

The relationships between the results on the EGRA subtasks were analyzed for the correlations between the scores obtained by students on each of these subtasks. The possible correlation score ranges from 0 to 1. A correlation closer to 1 indicates a stronger relationship between how students performed on that particular subtask and how they performed on the EGRA overall. A correlation closer to 0 indicates a weaker relationship between how students performed on that particular subtask and how they performed on the EGRA overall.

All of the correlations from the Midline 1 EGRA are statistically significant ($p<0.000$). Just as was observed at Baseline, the correlations are all positive, indicating that the students who scored the highest on one subtask are those who scored the highest on the other subtasks as well (see Table 16). Additionally, the correlations were particularly high between the subtasks of syllable identification, nonword reading and passage reading. Among all the EGRA subtasks, the listening comprehension subtask showed the lowest correlation with the other subtasks. The observations are also similar to what was observed at Baseline.

TABLE 16. CORRELATION MATRIX BETWEEN THE NUMBER OF CORRECT ANSWERS FOR EACH OF THE EGRA SUBTASKS

	Phonemic Awareness	Syllable Identification	Nonword Reading	Passage Reading	Reading Comprehension	Listening Comprehension
Phonemic Awareness	1					
Syllable Identification	0.603	1				
Nonword Reading	0.586	**0.911**	1			
Passage Reading	0.566	**0.892**	**0.919**	1		
Reading Comprehension	0.346	0.522	0.545	0.656	1	
Listening Comprehension	0.305	0.341	0.308	0.349	0.411	1

CHAPTER 4: CONTEXTUAL VARIABLES AND EGRA PERFORMANCE

4.1 Student Questionnaire

The student questionnaire dealt with familial and academic characteristics that could potentially have an influence on student learning.[11] Analyses were conducted to examine these variables in relationship to students' progress since Baseline on the EGRA subtasks. The aim is to identify which factors are linked to student progress in acquiring reading skills. Analyses were conducted to explore the relationship between all variables and EGRA performance. The results presented below are only for those variables which were found to have a statistically significant correlation.

Sex of the student

For three of the EGRA subtasks, male students had a greater increase in mean score since Baseline compared to female students. On the syllable identification subtask, male students, on average, correctly identified 2.3 syllables more than female students. As for the nonword reading subtask, male students read, on average, 0.79 more nonwords than female students. Finally, in the passage reading subtask, male students read an average of 1.8 more words than the female students. For these three subtasks, the difference in progress by sex between Baseline and Midline 1 was similar in both the control and experimental groups (see Table 17).

TABLE 17. PROGRESS IN PERFORMANCE BY SUBTASK ACCORDING TO THE SEX OF THE STUDENT AND THE GROUP

	Total				Control Group				Experimental Group			
	Girls		**Boys**		**Girls**		**Boys**		**Girls**		**Boys**	
	Mean	**SD**	**Mean**	**SD**	**Mean**	**SD**	**Mean**	**SD**	**Mean**	**SD**	**Mean**	**SD**
Phonemic Awareness	2.21	4.33	2.44	4.54	1.65	4.26	1.47	4.32	2.79	4.34	3.39	4.55
Syllable Identification*	13.37	13.51	14.99	14.60	12.86	13.15	15.19	14.77	13.89	13.88	14.79	14.44
Nonword Reading*	4.17	5.38	4.72	5.94	4.27	5.14	5.10	6.15	4.07	5.62	4.35	5.71
Passage Reading*	6.87	6.97	8.08	7.37	6.55	6.67	8.45	7.60	7.19	7.25	7.71	7.12
Reading Comprehension	-0.05	0.58	-0.03	0.64	-0.03	0.61	-0.01	0.69	-0.06	0.55	-0.04	0.59
Listening Comprehension	0.52	1.31	0.53	1.28	0.41	1.31	0.46	1.26	0.63	1.31	0.60	1.29

*Statistically significant difference ($p<0.05$)

[11] Descriptive analyses of the student questionnaire data are not presented in this report. Descriptive statistics of the Baseline student questionnaire data were presented in "USAID/Morocco: Reading for Success – Small-Scale Experimentation Activity (RFS-SSE) Early Grade Reading Assessment Baseline Report" prepared by School-to-School International and Chemonics International, Inc. in May 2016. Because the Midline 1 was administered only a few months after the Baseline, the characteristics studied by this questionnaire were not likely to have changed.

Language spoken at home

The progress of students since Baseline varied as a function of the language spoken at home and to which study group the student belonged. The majority of students surveyed speak Darija at home (n=1,274, 75.9%) while only a quarter of the students surveyed speak Amazigh (n=403, 24.1%). This variation is statistically significant for three subtasks. For the phonemic awareness subtask, students from the control group speaking Amazigh at home showed greater progress since Baseline than those students speaking Darija. For the experimental group, the reverse was observed: students speaking Darija at home showed greater progress. For the syllable identification and nonword reading subtasks, the students from the control group speaking Amazigh at home showed greater progress, while in the experimental group, the progress was similar for both groups (see Table 18).

TABLE 18. PROGRESS IN PERFORMANCE BY SUBTASK ACCORDING TO THE LANGUAGE SPOKEN AT HOME AND THE GROUP

	Total				Control Group				Experimental Group			
	Darija		**Amazigh**		**Darija**		**Amazigh**		**Darija**		**Amazigh**	
	Mean	**SD**	**Mean**	**SD**	**Mean**	**SD**	**Mean**	**SD**	**Mean**	**SD**	**Mean**	**SD**
Phonemic Awareness	2.30	4.46	2.39	4.37	1.42*	4.24	2.02*	4.42	3.18*	4.49	2.77*	4.29
Syllable Identification*	13.55	13.88	16.02	14.49	12.69	13.47	17.92	14.79	14.40	14.22	14.08	13.95
Nonword Reading*	4.21	5.45	5.15	6.22	4.19	5.24	6.15	6.59	4.23	5.67	4.14	5.65
Passage Reading	7.06*	7.05	8.67*	7.49	6.80	6.89	9.44	7.71	7.31	7.19	7.87	7.18
Reading Comprehension	-0.03	0.62	-0.05	0.60	-0.02	0.64	-0.04	0.65	-0.05	0.59	-0.07	0.54
Listening Comprehension	0.50	1.32	0.61	1.22	0.40	1.31	0.55	1.22	0.60	1.32	0.66	1.23

*Statistically significant difference ($p<0.05$)

Repeating of the class

There was no difference observed on EGRA performance since baseline between students who repeated first grade and those who did not, with the exception of two subtasks. For the nonword reading subtask, students who were repeating Grade 1 showed less progress (on average, 2.3 words less) compared to the students who were not repeating. For the passage reading subtask, the average progress of students who were repeating Grade 1 was 1.5 fewer words (see

Table 19).

TABLE 19. PROGRESS IN PERFORMANCE BY SUBTASK ACCORDING TO WHETHER THE STUDENT WAS REPEATING GRADE 1 AND THEIR STUDY GROUP

	Total				Control group				Experimental group			
	Non-repeating		**Repeating**		**Non-repeating**		**Repeating**		**Non-repeating**		**Repeating**	
	Mean	**SD**	**Mean**	**SD**	**Mean**	**SD**	**Mean**	**SD**	**Mean**	**SD**	**Mean**	**SD**
Phonemic Awareness	2.36	4.48	2.04	4.04	1.56	4.34	1.60	3.87	3.18	4.48	2.41	4.16
Syllable Identification	14.38	14.19	12.30	12.87	14.13	14.21	12.43	11.72	14.63	14.18	12.20	13.81
Nonword Reading*	4.54	5.72	3.59	5.01	4.79	5.75	3.56	4.56	4.30	5.69	3.61	5.38
Passage Reading*	7.70	7.23	5.48	6.56	7.73	7.23	4.99	6.31	7.67	7.23	5.89	6.75
Reading Comprehension	-0.04	0.63	-0.02	0.42	-0.03	0.67	0.01	0.29	-0.05	0.58	-0.05	0.52
Listening Comprehension	0.54	1.32	0.44	1.12	0.44	1.30	0.40	1.19	0.64	1.32	0.48	1.07

*Statistically significant difference between repeaters and non-repeaters ($p<0.05$)

Reaction of the teacher to incorrect answers from the student

For the passage reading subtask, students who reported that their teacher corrected errors in their exercise book, read, on average, 1.9 more words than the students who did not report that their teacher corrected errors in their exercise book (see Table 20). For all other subtasks, there was no statistically significant difference.

TABLE 20. PROGRESS IN THE PERFORMANCE BY SUBTASK ACCORDING TO THE REACTION OF THE TEACHER TO AN INCORRECT ANSWER AND THE GROUP

	Total				Control Group				Experimental Group			
	Other Reactions		**Corrected Errors**		**Other Reactions**		**Corrected Errors**		**Other Reactions**		**Corrected Errors**	
	Mean	**SD**	**Mean**	**SD**	**Mean**	**SD**	**Mean**	**SD**	**Mean**	**SD**	**Mean**	**SD**
Phonemic awareness	2.48	4.32	2.28	4.53	1.59	3.88	1.54	4.49	3.30	4.55	3.12	4.44
Syllable identification	13.12*	13.69	14.99*	14.25	13.19	13.54	14.39	14.19	13.06	13.84	15.66	14.31
Nonword reading	3.82*	5.44	4.82*	5.76	4.19	5.33	4.90	5.72	3.48	5.40	4.72	5.81
Passage reading*	6.17	6.79	8.18	7.28	5.92	6.53	8.19	7.32	6.39	7.02	8.16	7.25

Reading comprehension	-0.01	0.54	-0.05	.65	0.01	0.51	-0.04	0.17	-0.02	0.58	-0.06	0.59
Listening comprehension	0.50	1.26	0.54	1.33	0.39	1.23	0.46	1.32	0.59	1.29	0.64	1.33

*Statistically significant difference ($p<0.05$)

Status of literacy at home

For the passage reading subtask, students' progress varied according to the reading ability of the parents. No other subtasks showed a statistically significant difference between students who reported that their parents can read versus those who reported that their parents cannot read. On the passage reading subtask, students who said that neither parent knew how to read showed less progress than those whose father knew how to read or where both parents knew how to read. In addition, progress in the reading comprehension subtask also varied as a function of the reading ability of the parents (see Table 21).

TABLE 21. PROGRESS IN PERFORMANCE BY SUBTASK ACCORDING TO THE READING ABILITY OF THE PARENTS AND THE GROUP

	Control Group							
	Neither Parent		**Father Only**		**Mother Only**		**Both Parents**	
	Mean Score	**Standard Deviation**	**Mean Score**	**Standard Deviation**	**Mean Score**	**Standard Deviation**	**Mean Score**	**Standard Deviation**
Phonemic Awareness	1.45	4.25	1.76	4.26	0.41	3.96	1.76	4.36
Syllable Identification	14.91	14.85	14.80	15.23	12.66	10.93	13.47	13.61
Nonword Reading	5.07	5.79	4.90	5.91	4.37	5.85	4.45	5.44
Passage Reading*	5.66	7.18	8.38	7.96	6.55	6.72	7.92	6.79
Reading Comprehension	0.09	0.43	-0.07	0.63	-0.14	0.59	-0.02	0.72
Oral Comprehension	0.45	1.12	0.43	1.31	0.30	1.36	0.46	1.33
	Experimental Group							
	Neither Parent		**Father Only**		**Mother Only**		**Both Parents**	
	Mean Score	**Standard Deviation**	**Mean Score**	**Standard Deviation**	**Mean Score**	**Standard Deviation**	**Mean Score**	**Standard Deviation**
Phonemic Awareness	2.78	4.36	3.25	4.34	2.77	4.74	3.19	4.46
Syllable Identification	12.87	13.41	15.77	15.52	12.74	13.50	14.62	13.99
Nonword Reading	3.88	4.66	4.46	6.52	4.14	5.98	4.24	5.59
Passage Reading*	5.88	6.86	7.75	8.00	6.79	6.63	8.00	7.02
Reading Comprehension	0.02	0.43	0.00	0.56	-0.07	0.53	-0.09	0.63
Oral Comprehension	0.42	1.06	0.75	1.22	0.59	1.25	0.64	1.40

*Statistically significant difference (p<0.05)

As for the presence of books, magazines or newspapers in the home, the performance of students varied only in the ability to read a passage. Students who said they had these types of documents at home read an average of 1.6 words more than those who did not have these documents at home (see Table 22).

TABLE 22. PROGRESS IN PERFORMANCE BY SUBTASK ACCORDING TO WHETHER THE STUDENT HAD BOOKS, MAGAZINES OR NEWSPAPERS AT HOME AND THE GROUP

	Control Group				Experimental Group			
	No		Yes		No		Yes	
	Mean Score	Standard Deviation	Mean Score	Standard Deviation	Mean Score	Standard Deviation	Mean Score	Standard Deviation
Phonemic Awareness	1.52	4.17	1.66	4.54	3.19	4.55	2.93	4.31
Syllable Identification	13.45	13.83	15.05	14.26	14.29	13.93	14.44	14.55
Nonword Reading	4.62	5.69	4.75	5.59	4.24	5.39	4.17	6.07
Passage Reading*	6.75	7.20	8.93	6.93	6.79	7.03	8.45	7.34
Reading Comprehension	-0.01	0.55	-0.06	0.81	-0.01	0.50	-0.12	0.67
Listening Comprehension	0.44	1.20	0.44	1.46	0.56	1.29	0.71	1.32

*Statistically significant difference (p<0.05)

4.2 Teacher Questionnaire

In total, 87 teachers provided answers to this questionnaire at Midline 1. Three teachers were absent at the time of data collection. The results presented in this section examine the relationships between the various variables of the teacher's questionnaire and the progress of student performance on the EGRA between baseline and Midline 1. Only results that are statistically significant are presented. See ANNEX C: SSME Tools for a complete listing of variables included in the Teacher Questionnaire.

Characteristics of the teachers

Students' improvement on the Phonemic Awareness subtask varied according to the sex of the teacher. For the control group, students taught by female teachers improved by an average of 0.9 correct answers over students taught by male teachers. For the experimental group, it was students taught by male teachers who demonstrated a greater improvement compared students taught by female teachers, who experienced an average of 0.4 more correct answers (see Table 23).

TABLE 23. PROGRESS IN THE PERFORMANCE BY SUBTASK ACCORDING TO THE SEX OF THE TEACHER AND THE GROUP

	Total				Control Group				Experimental Group			
	Female		Male		Female		Male		Female		Male	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD

Phonemic Awareness	2.51	1.50	2.23	1.62	2.06*	1.47	1.22*	1.05	2.88*	1.46	3.27*	1.45
Syllable Identification	14.58	5.44	14.14	6.46	14.57	5.97	14.24	7.26	14.57	5.21	14.04	5.68
Nonword Reading	4.51	2.50	4.48	2.30	4.74	2.85	4.83	2.44	4.31	2.22	4.11	2.14
Passage Reading	8.57*	2.99	6.53*	3.33	8.69	3.12	6.80	3.58	8.47	2.94	6.25	3.11
Reading Comprehension	-0.07	0.22	-0.02	0.16	-0.04	0.26	-0.03	0.19	-0.09	0.19	-0.01	0.12
Listening Comprehension	0.59	0.33	0.48	0.41	0.49	0.27	0.39	0.41	0.66	0.36	0.56	0.40

*Statistically significant difference between girls and boys ($p<0.05$)

The number of students in a class is correlated with improved performance on the Syllable Identification, Nonword Reading, and Passage Reading subtasks. For these three subtasks, the correlations are statistically significant, which indicates that the larger the class size, the less students' performance improved on the subtasks. These correlations are similar for both the experimental and control groups, with the exception of the Passage Reading subtask which showed a weaker correlation for the experimental group (see Table 24).

TABLE 24. CORRELATION BETWEEN THE NUMBER OF STUDENTS IN THE CLASS AND THE PROGRESS IN PERFORMANCE BY SUBTASK

	Control Group	**Experimental Group**
Phonemic Awareness	-0.12	-0.02
Syllable Identification	-0.34*	-0.34*
Nonword Reading	-0.38*	-0.31*
Passage Reading	-0.35*	-0.19*
Reading Comprehension	-0.15	-0.14
Listening Comprehension	-0.11	-0.23

*Statistically significant correlation ($p<0.05$)

For the control group, the number of students in class who were repeating the grade was related to improved performance on the Passage Reading subtask (see Table 25). Because this correlation was statistically significant, it indicates that the more repeating students in a class, the less elevated student improvement on the subtask was. This same correlation was not statistically significant for the experimental group. For this group, it was the Reading Comprehension subtask that presented a statistically significant correlation with the number of repeating students. Because this correlation was negative, it indicates that a greater number of repeating students in a class is associated with a smaller improvement in reading comprehension.

TABLE 25. CORRELATION BETWEEN THE NUMBER OF REPEATING STUDENTS IN THE CLASS AND THE PROGRESS IN PERFORMANCE BY SUBTASK

	Control Group	**Experimental Group**
Phonemic Awareness	-0.32	-0.14
Syllable Identification	0.05	-0.07
Nonword Reading	-0.01	-0.14
Passage Reading	-0.24*	0.09

Reading Comprehension	0.12	-0.32*
Listening Comprehension	-0.01	0.09

*Statistically significant correlation (p<0.05)

4.3 School Director Questionnaires

Out of 90 directors who answered the questionnaire at baseline, 87 answered the questionnaire again at Midline 1. As with the student and teacher questionnaires, the relationships between characteristics explored in the school director questionnaire and the progress of student performance were examined. The results are presented below for the only variable from the school director questionnaire that was found to have a statistically significant relationship to students' progress on the EGRA since Baseline. See ANNEX C: SSME Tools for a complete listing of variables included in the School Director Questionnaire.

Personal characteristics of the directors

A school director's number of years of experience is related to progress on the performance of students for certain subtasks. For the control group, a positive correlation is observed for the passage reading subtask only: The more years of experience the director had, the more the students from his/her school showed progress, on average, on the reading passage subtask. For the experimental group, the relationship is statistically significant for the subtasks of phonemic awareness, syllable identification, and nonword reading. For all three of these subtasks, the correlation is negative: The more years of experience the director had, the less the students from his/her school showed progress, on average (see Table 26).

TABLE 26. CORRELATION BETWEEN THE NUMBER OF YEARS OF EXPERIENCE OF THE DIRECTOR AND THE PROGRESS IN PERFORMANCE BY SUBTASK

	Control Group	**Experimental Group**
Phonemes	0.273	-0.353*
Syllables	0.194	-0.348*
Nonwords	0.078	-0.331*
Reading of A Passage	0.310*	-0.097
Reading Comprehension	-0.257	-0.147
Listening Comprehension	0.193	0.983

*Statistically significant correlation (p<0.05)

4.4 Classroom Inventory Chart

The data from the classroom inventory chart is available for 87 of 90 schools that participated in the Midline 1. Just as in the other sections of the report, the results are shown only for the variables that had a statistically significant relationship with the progress of the students on the various EGRA subtasks. See ANNEX C: SSME Tools for a complete listing of variables included in the Classroom Inventory Chart.

For the experimental group, the number of classes currently functioning at the time of observation and improved reading comprehension scores presented a statistically significant correlation (see

Table 27). Because this correlation was negative, it indicates that the more classes that were functioning in a school, the weaker student reading comprehension performance.

TABLE 27. CORRELATION BETWEEN THE NUMBER OF CLASSES CURRENTLY FUNCTIONING AND THE PROGRESS IN PERFORMANCE BY SUBTASK

	Control Group	Experimental Group
Phonemes	0.023	-0.134
Syllables	-0.244	-0.075
Nonwords	-0.277	-0.088
Reading of A Passage	0.006	0.228
Reading Comprehension	-0.125	-0.321*
Listening Comprehension	-0.085	-0.122

*Statistically significant correlation ($p<0.05$)

Finally, a statistically significant correlation was observed in the experimental group between the number of students in class at the time of observation and the improvement of student performance on the Syllable Identification subtask (see Table 28). Because this correlation was statistically significant, it shows that the more students there were in a class, the less improvement students showed.

TABLE 28. CORRELATION BETWEEN THE NUMBER OF STUDENTS IN THE CLASS AND THE PROGRESS IN PERFORMANCE BY SUBTASK

	Control Group	Experimental Group
Phonemes	-0.111	0.031
Syllable	-0.166	-0.297*
Nonwords	-0.230	-0.288
Reading of A Passage	-0.054	-0.139
Reading Comprehension	-0.136	-0.150
Listening Comprehension	0.019	-0.238

*Statistically significant correlation ($p<0.05$)

4.5 Classroom Observation

At Midline 1, classroom observations were conducted on the same 89 teachers who were surveyed. The observations lasted 30 minutes and took place during a reading lesson. The classroom observation captured the presence of five key activities at three-minute intervals throughout the lesson observation. For example, the first item noted what teachers were focusing on: the whole classroom, only one student, a small group, something else/not students, or not present in the classroom. By looking at how often each item was observed, we can estimate how much time teachers spent on each activity throughout the observation. See ANNEX C: SSME Tools for the full classroom observation tools.

In both the experimental and control groups, teachers spent the majority of the time focusing on the entire classroom, as opposed to individual students, small groups, or something else (see

Table 29).

TABLE 29. PRIMARY FOCUS OF TEACHER

The teacher gives his attention to…	Midline 1		Baseline	
	Control Group	**Exp. Group**	**Control Group**	**Exp. Group**
The whole classroom**	58.10%*	76.04%*	68.57%	75.81%
Only one student**	33.33%*	20.00%*	23.81%	19.53%
Something else/does not concentrate on students	3.10%	1.63%	2.38%	2.09%
A small group	2.62%	1.86%	3.09%	1.63%
The teacher is not present in the classroom	0.00%	0.00%	0.00%	0.00%

*Statistically significant difference between study groups at one time point ($p<.05$).
**Statistically significant difference ($p<.05$) at Midline 1 after controlling for differences at Baseline.

Additional analyses were done to examine the difference between the two groups of teachers (control and experimental) at Midline 1 while controlling for any differences in the two groups observed at Baseline. After controlling for the differences between the groups at Baseline, a statistically significant difference was found between the groups at Midline 1 for the time teachers spent giving attention to the whole classroom and doing something else/not concentrating on students. Results suggest that teachers in the experimental group spend about 15.00% more time giving attention to the whole classroom than teachers in the control group. Results also suggest that teachers in the experimental group spend about 11.47% less time focusing on only one student compared to control.

Ten pedagogic content areas were captured during the classroom observation. To understand what areas of reading skills teachers are focusing on, these ten activities were grouped into five broader skill areas as follows:

1. Pre-reading skills: non-written sounds, letters/sounds
2. Reading skills: reading of vocabulary, reading of sentences, vocabulary-meaning of words
3. Writing skills: writing/dictation, writing-creation of texts
4. Reading and comprehending connected text skills: reading of text, reading and comprehension of text
5. Other

At Midline 1, teachers from the experimental group were found to spend more time on pre-reading skills than teachers from the control group. Conversely, teachers from the control group were found to spend more time on reading and comprehending connected text skills than teachers from the experimental group (see Table 30).

TABLE 30. PRIMARY CONTENT OF LESSON

The pedagogic content of the lesson…	Midline 1		Baseline	
	Control Group	**Exp. Group**	**Control Group**	**Exp. Group**

Pre-reading skills**	7.14%*	32.80%*	24.05%	21.40%
Reading and comprehending connected text skills**	45.95%*	25.12%*	10.71%	15.35%
Reading skills	32.62%	24.42%	45.48%	45.35%
Writing skills	10.48%	14.65%	17.14%	15.12%
Other	.95%	1.76%	.48%	1.86%

*Statistically significant difference between study groups at one time point (p<.05).
**Statistically significant difference (p<.05) at Midline 1 after controlling for differences at Baseline.

After controlling for the differences between the groups at Baseline, a statistically significant difference was found between the groups at Midline 1 for the time teachers spent on pre-reading skills and on reading and comprehending connected text skills. Results suggest that teachers in the experimental group spend about 26.27% more time on pre-reading skills than teachers in the control group. Results also suggest that teachers in the experimental group spend about 21.91% less time on reading and comprehending connected text skills than teachers in the control group.

Teachers from both the experimental and control groups were observed spending the most time listening to and/or observing the students. It was also found that teachers in the experimental group spent more time asking questions to students compared to teachers from the control group (see Table 31).

TABLE 31. PRIMARY ACTIVITY OF TEACHER

Activities of the teacher…	Midline 1		Baseline	
	Control Group	**Exp. Group**	**Control Group**	**Exp. Group**
Listen to the students/ Observe the students**	62.14%*	45.81%*	52.62%	57.44%
Ask a question**	16.19%*	29.07%*	15.71%	19.53%
Explain/ Communicate orally	6.90%	10.47%	6.90%	6.51%
Read out loud	6.67%	7.67%	11.90%	9.78%
Write	5.24%	6.51%	10.71%*	5.81%*

*Statistically significant difference between study groups at one time point (p<.05).
**Statistically significant difference (p<.05) at Midline 1 after controlling for differences at Baseline.

After controlling for the differences between the groups at Baseline, a statistically significant difference was found between the groups at Midline 1 for the time teachers spent listening to or observing students and the time teachers spent on asking questions to the students. Results suggest that teachers in the experimental group spend about 14.96% less time listening to or observing students than teachers in the control group. Results also suggest that teachers in the experimental group spend about 13.05% more time asking students questions than teachers in the control group.

The classroom observation also noted what students were doing throughout the lesson. Eleven student activities were captured during the classroom observation. To understand what type of activities students spend their time on, these eleven activities were grouped into four broader categories as follows:

1. Single student-focused activities: individually read out loud, write on the board, respond to a question, ask a question
2. Whole class-focused activities: listen to/watch the teacher, read/read back, read collectively
3. Independent activities: read silently, write on paper individually
4. Non-academic activities: other (play, etc.), lack of concentration (speaking, sleeping, playing)

At Midline 1, students in both groups spent the majority of time throughout the lesson on single student-focused activities. Whole class-focused activities were the next most common type of activity in both groups (see

Table 32).

TABLE 32. PRIMARY ACTIVITY OF STUDENTS

Activities of the students…	Midline 1		Baseline	
	Control Group	**Exp. Group**	**Control Group**	**Exp. Group**
Single student-focused activities**	70.24%*	61.16%*	63.57%	69.07%
Whole class-focused activities**	15.48%*	26.51%*	18.33%	13.95%
Independent activities	10.95%	9.53%	13.33%	13.95%
Non-academic activities**	.48%*	2.33%*	2.62%	2.09%

*Statistically significant difference ($p<.05$).

**Statistically significant difference ($p<.05$) at Midline 1 after controlling for differences at Baseline.

After controlling for the differences between the groups at Baseline, a statistically significant difference was found between the groups at Midline 1 for the time students spent on whole class-focused activities, on single student-focused activities, and on non-academic activities. Results suggest that students in the experimental group spend about 11.57% more time on whole class-focused activities and 9.45% less time on single student-focused activities than students in the control group. Results also suggest that students in the experimental group spend about 1.89% more time on non-academic activities than students in the control group.

Finally, the observation noted which pedagogic tools teachers were using such as textbooks, posters, handouts, etc. A full list of the tools included in the observation form can be found in ANNEX C: SSME Tools. At Midline 1, teachers in the experimental group were found to use .63 more tools than teachers in the control group ($p=.01$).

4.6 Parent Questionnaire

At Midline 1, 173 parents from the 90 schools (up to two parents per school) were surveyed. Parents at each school were selected to participate in the survey based on convenience. The parents surveyed are not necessarily parents of the students who participated in the EGRA; therefore, conclusions cannot be drawn about the relationship between the parent survey responses and EGRA performance. Rather, the parent survey was designed to provide contextual background information only about the home environment. The RFS-SSE program includes no intervention targeted towards parents.

Because a new group of parents were surveyed at Midline 1 compared to baseline, and the results cannot be related to EGRA performance, descriptive analyses only of the Midline 1 survey responses are presented here.

Of the parents who participated in the survey, 57.0% (n=98) were male and 43.0% (n=74) were female. On average, students took 12.24 minutes to get to school. The time varied between 1 and 60 minutes (see Table 33).

TABLE 33. NUMBER OF MINUTES REQUIRED TO GET TO SCHOOL

	n	Average	Standard deviation	Range
Number of minutes	172	13.1	12.0	1 – 60

Fifty-five percent (55.49%, n=96) of the parents reported that the school had a parents' association. It is to be noted that about one in five parents (22.54%, n=39) said that they were unaware if the school had a parents' association. Among the parents who said that the school had a parents' association, 63.24% (n=43) had the impression that the association contributed to the improvement of the school.

The majority of parents (98.8%, n=169) said that they receive their child's grades. The parents reported that their child was absent from school primarily for reasons of illness (89.53%, n=154) (see Table 34). Finally, 28.32% (n=49) of the parents reported that their child had been the victim of violence during the year.[12]

TABLE 34. REASON FOR STUDENTS' ABSENTEEISM

	n (%)
Illness	154 (89.53%)
Other	14 (8.14%)
Weather conditions	11 (11.05%)
Doesn't want to go to school	7 (4.07%)
Does not get up early enough in the morning	6 (3.49%)
Funeral	5 (2.91%)
Market day/preparation	2 (1.16%)
School uniform not ready	2 (1.16%)
Taking care of a sibling	1 (0.58%)
Taking care of an ill family member	1 (0.58%)
Other household chores	1 (0.58%)
Problems related to transportation	1 (0.58%)
Lack of food	0 (0%)

[12] Note that this figure is higher than it was at Baseline. At Baseline, only 8.8% (n=15 of 172) of parents responded that their child had been the victim of violence.

CHAPTER 5: CONCLUSION AND RECOMMENDATIONS

A summary of the Midline 1 results by subtask with a focus on the key takeaway message is provided below, followed by recommendations. As there was only two and a half months of intervention implementation between the Baseline and Midline 1, conclusions that can be drawn about the effectiveness of the intervention are limited.

That being said, the Midline 1 evaluation does provide some initial indications which can be confirmed through further evaluation at the Midline 2 and Endline.

- **Phonemic Awareness**
  The results indicate a positive impact for students exposed to the interventions to build phonemic awareness skills. The development of phonemic awareness skills in young children is an important building block to later success with fluent reading. Some students are naturally more receptive to the development of this skill, while others require a more directive approach.
- **Syllable identification**
  The results indicated no differences in syllable identification for students exposed to the interventions versus students in the control group. Similar to phonemic awareness, the ability for young children to identify syllables is predictive of future reading success. Students have to be taught to hear the demarcations of the syllables in a direct, explicit, engaging manner. The syllable knowledge will transfer to reading, spelling, and writing, as the child develops further academic skills.
- **Non-word Reading**
  The results indicated no differences in non-word reading for students exposed to the treatment versus students in the control group. Non-word reading is an indicator of phonics skill, the ability to decode words.
- **Passage reading**
  The results indicated no differences in passage reading for students exposed to the treatment versus students in the control group. Before students can "read to learn", they must be able to read fluently and with comprehension.
- **Reading Comprehension and Listening Comprehension**
  The results indicated no differences in reading comprehension for students exposed to the treatment versus students in the control group, whereas students exposed to the interventions had higher listening comprehension skills than the control students.

As the intervention began during the second semester of Grade 1, RFS-SSE encouraged feedback from the terrain in the forms of questionnaires and focus groups over the course of the semester. The following recommendations include suggestions that RFS-SSE is addressing by revising materials prior to the start of the new academic year in September 2016 or intends to implement during the course of the upcoming academic year. By following these recommendations, RFS-SSE will be able to implement the approach and materials as designed and evaluate more clearly the success of the intervention at the Endline in May 2017.

- Adopt the lesson structure to student learning pace - increasing the difficulty of the lessons and the lengths of texts during the course of the year.
- Increase the number of days for teacher training both during the initial training and during the course of the academic year.
- Emphasize practical training rather than the theoretical approach during both the training of trainer's and teachers workshops.
- Increase time allocated for the phonemics awareness track per daily class in Grade 1.
- Provide coaching and teacher support throughout the year.
- Develop a student exercise book to be used by the students to accompany the activities laid out in the teacher's guidebook.

- Encourage school administration and teachers to further engage with the parents to ensure that they are aware and, if possible, understand the objectives of the new reading method.
- Provide teachers with pedagogical tools that support the approach and the individual lesson plans.
- Encourage enhanced community support for literacy through civil society efforts such as summer camp and summer reading activities.

ANNEXES

ANNEX A: Reliability of the EGRA Subtask Items

Subtask: Phonemic Awareness

Table 1. Indices of difficulty and discrimination for the phonemic awareness subtask

Variable	p	D
Item1	.63	0.73
Item2	.52	0.85
Item3	.58	0.78
Item4	.56	0.89
Item5	.59	0.91
Item6	.39	0.69
Item7	.57	0.91
Item8	.58	0.92
Item9	.54	0.91
Item10	.57	0.92

Subtask: Syllable identification

Table 2. Indices of difficulty and discrimination for syllable identification

Variable	p	d
Item1	.80	0.63
Item2	.87	0.48
Item3	.66	0.50
Item4	.78	0.65
Item5	.76	0.71
Item6	.73	0.70
Item7	.67	0.67
Item8	.78	0.65
Item9	.79	0.63
Item10	.68	0.72
Item11	.73	0.72
Item12	.70	0.72
Item13	.63	0.71
Item14	.71	0.72
Item15	.73	0.69
Item16	.56	0.68
Item17	.74	0.70
Item18	.76	0.69
Item19	.67	0.73
Item20	.75	0.70
Item21	.65	0.64
Item22	.55	0.65
Item23	.66	0.61
Item24	.45	0.55
Item25	.54	0.66
Item26	.60	0.70

Item27	.57	0.70
Item28	.59	0.71
Item29	.60	0.69
Item30	.56	0.73
Item31	.61	0.76
Item32	.58	0.77
Item33	.58	0.78
Item34	.57	0.78
Item35	.58	0.77
Item36	.57	0.78
Item37	.55	0.79
Item38	.55	0.78
Item39	.48	0.74
Item40	.52	0.79
Item41	.53	0.79
Item42	.51	0.80
Item43	.45	0.77
Item44	.49	0.80
Item45	.47	0.79
Item46	.44	0.76
Item47	.44	0.79
Item48	.43	0.78
Item49	.42	0.79
Item50	.41	0.78
Item51	.35	0.70
Item52	.35	0.73
Item53	.30	0.67
Item54	.32	0.70
Item55	.32	0.71
Item56	.32	0.72
Item57	.31	0.71
Item58	.29	0.70
Item59	.28	0.69
Item60	.24	0.64
Item61	.25	0.68
Item62	.23	0.65
Item63	.23	0.66
Item64	.21	0.63
Item65	.17	0.57
Item66	.18	0.60
Item67	.17	0.60
Item68	.15	0.57
Item69	.16	0.58
Item70	.15	0.56
Item71	.12	0.51
Item72	.10	0.47
Item73	.10	0.48
Item74	.08	0.43
Item75	.09	0.46
Item76	.08	0.45
Item77	.07	0.43
Item78	.06	0.42
Item79	.06	0.42

Item80	.06	0.41
Item81	.05	0.37
Item82	.06	0.39
Item83	.05	0.39
Item84	.05	0.38
Item85	.04	0.36
Item86	.04	0.35
Item87	.03	0.34
Item88	.03	0.32
Item89	.03	0.31
Item90	.03	0.30
Item91	.03	0.30
Item92	.02	0.29
Item93	.02	0.28
Item94	.02	0.26
Item95	.02	0.24
Item96	.01	0.22
Item97	.01	0.22
Item98	.01	0.21
Item99	.01	0.20
Item100	.01	0.16

Subtask: Nonword reading

Table 3. Indices of difficulty and discrimination for the reading of nonwords

Variable	p	d
Item1	.60	0.69
Item2	.62	0.68
Item3	.64	0.67
Item4	.61	0.74
Item5	.62	0.65
Item6	.55	0.63
Item7	.56	0.69
Item8	.56	0.68
Item9	.52	0.69
Item10	.42	0.58
Item11	.41	0.59
Item12	.40	0.64
Item13	.37	0.61
Item14	.41	0.72
Item15	.38	0.70
Item16	.36	0.71
Item17	.40	0.76
Item18	.35	0.74
Item19	.32	0.73
Item20	.29	0.71
Item21	.23	0.62
Item22	.24	0.68
Item23	.21	0.64
Item24	.21	0.64
Item25	.17	0.61

Item26	.14	0.55
Item27	.14	0.56
Item28	.11	0.51
Item29	.08	0.46
Item30	.07	0.44
Item31	.06	0.39
Item32	.04	0.34
Item33	.03	0.32
Item34	.02	0.30
Item35	.02	0.28
Item36	.02	0.25
Item37	.01	0.22
Item38	.01	0.20
Item39	.01	0.18
Item40	.01	0.16
Item41	.01	0.15
Item42	.01	0.15
Item43	.01	0.13
Item44	.01	0.11
Item45	.01	0.09
Item46	.01	0.09
Item47	.01	0.05
Item48	.01	0.06
Item49	.01	0.06
Item50	.01	0.03

Subtask: Reading Comprehension

Table 4. Indices of difficulties and discrimination for reading comprehension

Variable	p	d
Item1	.02	0.21
Item2	.15	0.47
Item3	.15	0.45
Item4	.01	0.24
Item5	.01	0.10

Subtask: Listening Comprehension

Table 5. Indices of difficulty and discrimination for listening comprehension

Variable	p	d
Item1	.44	0.29
Item2	.20	0.47
Item3	.24	0.47
Item4	.53	0.52
Item5	.26	0.56

ANNEX B: Enumerator Consistency During Training

Task	Consistency Measure 1	Consistency Measure 2
Phonemic Awareness	96.67%	98.67%
Syllable Identification	96%	96%
Nonword Reading	99%	99%
Passage Reading	96%	100%
Reading Comprehension	100%	100%
Listening Comprehension	98.67%	99.33%

ANNEX C: SSME Tools

(Chemonics Drive)